

Coexpression and cis-motif enrichment analyses

 Sandip M Kale  Ravi Koppolu

Updated date: May 22, 2021



An abbreviated version of this protocol was published in Science Advances in Apr 2021

Transcriptional landscapes of floral meristems in barley

DOI: 10.1126/sciadv.abf0832

Detailed protocol

Co-expression and cis-motif enrichment analysis:

Sandip Kale, Ravi Koppolu

Correspondance to kale@ipk-gatersleben.de, koppolu@ipk-gatersleben.de, schnurbusch@ipk-gatersleben.de

Gene filtration:

From the transcriptome data, TPM values for ~80K genes were obtained. The gene filtration was carried out for two reason:

1. To remove genes with low expression in most of the tissue/stage
2. To remove genes showing large variations within replicates

Two step procedure was used:

1. Calculate co-efficient of variation (CV) within replicate for each gene
2. Discard genes having CV above 25% in more than 50% of samples. Eg. We had 10 tissues and so we discarded the genes showing more than 25% variation in 5 or more tissues.

Co-expression analysis

The filtered genes were used for co-expression analysis. The network was constructed using R package WGCNA (Langfelder and Horvath 2008). Following parameter were considered for WGCNA

1. Pearson correlation matrix was calculated between all pairs of selected genes from which the adjacency matrix was calculated using signed network with the soft threshold of 18.
2. Average linkage hierarchical clustering was conducted for grouping of genes with highly similar co-expression patterns.
3. Modules were detected using Dynamic Hybrid Treecut (Langfelder et al. 2007) function using minimum module size of 50
4. The expression profile of each module was summarized by module eigen-gene defined as its first principal component, and modules with highly correlated (>0.75) eigen-genes were merged. Furthermore, modules with expression patterns similar to genes of interest were selected.

Identification of Hub genes:

The hub genes from each module were determined based on two statistics: 1) Degree: number of connections of a node (gene) 2) betweenness centrality: the number of times a path passes through the node (gene). The degree and betweenness values were calculated for each gene from a module using igraph package (Csardi and Nepusz 2006). The genes were ranked based on these two statistics and top 10 genes were considered as hub genes. The data was visualized using cystoscape 3.7.1 (Shannon et al. 2003).

Cis-motif enrichment analysis

The cis-motif analysis was carried out separately for genes from each module.

1. First, 1500bp upstream sequence for each gene was extracted from genome assembly using "bedtools flank" command.
2. We then used "findMotif.pl" program from HOMER package for simultaneous motif identification and study their enrichment (Heinz et al. 2010).
3. The program identifies known and novel motifs and perform enrichment against background sequences
4. The program was run to identify motifs of 4- to 10-bp sizes.
5. The enrichment of these motifs was searched against those calculated from background sequences defined by HOMER.

References

Csardi G, Nepusz T (2006) The igraph software package for complex network research. InterJournal, Complex Systems 1695. <https://igraph.org>
Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell 38:576-589
Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9:559
Langfelder P, Zhang B, Horvath S (2007) Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. Bioinformatics 24:719-720
Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Research 13:2498-2504

How to cite:(Readers should cite both the Bio-protocol preprint and the original research article where this protocol was used)

1. Kale, S. M. and Koppolu, R. (2021). Coexpression and cis-motif enrichment analyses. Bio-protocol Preprint. bio-protocol.org/preprint1103.
2. Thiel, J., Koppolu, R., Trautewig, C., Hertig, C., Kale, S. M., Erbe, S., Mascher, M., Himmelbach, A., Rutten, T., Esteban, E., Pasha, A., Kumlehn, J., Provart, N. J., Vanderauwera, S., Froberg, C. and Schnurbusch, T.(2021). Transcriptional landscapes of floral meristems in barley. Science Advances 7(18). DOI: [10.1126/sciadv.abf0832](https://doi.org/10.1126/sciadv.abf0832)

Copyright: Content may be subjected to copyright.